Preprints.org

Article

Not peer-reviewed version

An In-Depth Investigation into the Performance of State-of-the-Art Zero-Shot, Single-Shot, and Few-Shot Learning Approaches on an Out-of-Distribution Zero-Day Malware Attack Detection

Tosin Ige * , Christopher Kiekintveld , Aritran Piplai , Amy Wagler , Olukunle Kolade , Bolanle Hafiz Matti

Posted Date: 6 September 2024

doi: 10.20944/preprints202409.0509.v1

Keywords: Few-Shot Learning; One-Shot Learning; Zero-Shot Learning; Machine Learning; Deep Learning; Zero-Day; Malware

Article

An In-Depth Investigation into the Performance of State-of-the-Art Zero-Shot, Single-Shot, and Few-Shot Learning Approaches on an Out-of-Distribution Zero-Day Malware Attack Detection

Tosin Ige, Christopher Kiekintveld, Aritran Piplai, Amy Wagler, Olukunle Kolade and Bolanle Hafiz Matti

[1] Dept. of Computer Science, The University of Texas at El Paso, Texas, USA
[2] Dept. of Public Health Science, The University of Texas at El Paso, Texas, USA
[3] Office of Naval Research, United State Navy, Pentagon, USA
[4] Office of Network Security, Palo Alto Networks Inc, Texas, USA
* Correspondence: toige@miners.utep.edu

Abstract: N-shot learning has emerge in recent year as potential learning approach to solve the problem of data scarcity by learning underlying pattern from a few training sample. Despite recent state-of-the-art research on model-agnostic metal learning, transfer learning, and optimization strategy to rapidly learn valid information from few sample, there remains a big challenge on an actual out-of-distribution zero-day without any similarity to previously known malware family or new variant of an existing malware family. This ultimately questions the effectiveness of current state-of-the-art few-shot learning approach. In this research, we did an in-depth investigation into the performance of state-of-the-art Zero-shot, Single-shot, and few-shot learning approaches on zero-day out-of-distribution malware attack detection based on their static properties using Malimg and Malevis malware dataset. We ensure our model was aware of an out-of-distribution class during training while varying the number of samples in the out-of-distribution class accordingly zero-shot(no sample), single-shot (1 sample), few-shot(5 samples) while using confusion matrix to get the actual number of correct prediction on out-of-distribution malware validation samples. we assert that the model should be smart enough to detect and classify previously unseen data into an empty family as an out-of-distribution considering that the model was made to be aware of the existence of such distribution during training. Result shows 0, 0, and 3 correct out-of-distribution predictions on Zero-shot, single-shot, and few-shot experiments respectively, thereby showing limitation of the current state-of-the-art N-shot approaches on out-of-distribution attack.

Keywords: few-shot learning; one-shot learning; zero-shot learning; machine learning; deep learning, zero-day; malware

1. Introduction

Malware currently stands as the fastest-growing threat with 41% of enterprises witnessing a malware attack in just concluded year 2023 followed by phishing and ransomware attack. In year 2023 alone, the number of enterprises experiencing ransomware attacks increased by over 27% with only 8% of businesses attacked resorting to paying the ransom demands resulting in significant financial loss in addition to losses incurred due to downtime. There are 95 new families of malware in year 2022 alone averaging 1 new family every 4 days aside variants while year 2023 witness 43 new malware families averaging 1 new malware family per week aside variants making emerging malware families a major threat to cybersecurity causing damages worth billions of Dollars annually. The ease with which attacker creates new variants of malware coupled with the rate at which new variants are being release poses a real challenge both for their detection, identification and classification, reason being that machine learning and deep learning model are only effective in detecting previously seen variants during training. To identify malware, traditional signature-based analysis requires effort of an expert to generate hand-designed signatures which is highly impossible to achieve in light of the ease and frequency at which new variants quickly emerge by the simple use of polymorphic or metamorphic techniques. Machine learning had grown to become the mainstream trend for efficient malware identification and signature generation due to its ability to learn from relevant malware features [1] from dataset and to efficiently use knowledge obtained from training for accurate prediction. In particular, deep learning based models from Recurrent Neural Network, Long-Short Term Memory, and Convolusional Neural Network had proven to be effective in the prediction of malware [2,3]

Despite the effectiveness of machine learning and deep learning in the identification of malware, One major problem with the resulting model is that they are only good at predicting malware that are previously seen during training provided and on the condition that the data is large enough for the model to learn from, they often performed poor against data not previously seen during training, hence their vulnerability to Zero-Day malware or variants. The ease at which new malware variants could be developed coupled with the high frequency rate in which previously unseen malware are being released to the public makes the problem more potent by worsening it more as it will be practically impossible for the rate of re-training a model to be equal to the rate at which new malware variants are being release, even if the strategy of re-training model is adopted for every new malware variants, getting enough data of every new variants will be nearly impossible to get posing another mountain of challenge considering the fact that machine learning and deep learning model learns from data hence, their heavy reliability on data[4,5].

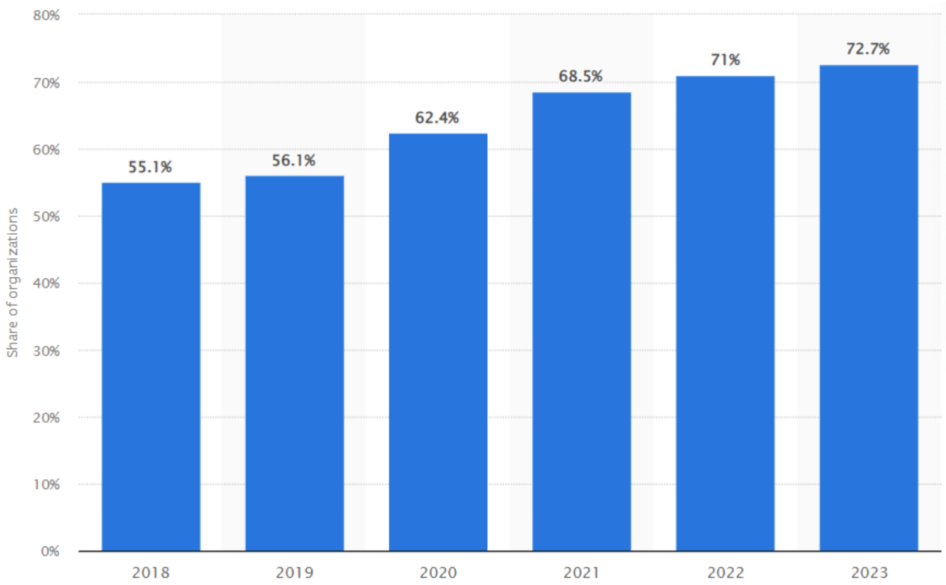

Figure 1. Annual share of organizations affected by ransomware attacks worldwide from 2018 to 2023 showing annual increase in the number of successful ransomware attack despite recent state-of-the-art research to few-shot learning

In order to resolve the problem of data scarcity for new malware variants in such a way that models will be able to learn from very few samples and improves its accuracy overtime, several state-of-the-art research had been carried out on few-short learning as a possible solution [6,7], but despite recent state-of-the-art research and development on few-shot learning approach as a cyber defense approaches to effectively protect computer system and critical infrastructure from malware, the ever increasing rate at which sophisticated variants of malware are created to exploit zero day vulnerabilities had continues to defy odds by making it difficult to classify previously unseen malware or its variants into correct families despite recent advancement in the state-of-the-art approach [8,9], the reason for the difficulty is being that the model had not previous seen the new variant during training thereby making the classification difficult to predict. Zero-day out-of-distribution prediction problem is exacerbated by sample scarcity due to challenges associated with the collection of a large volume of a newly detected variants or malware family to train a classifier which is extremely hard coupled with the unavoidability of over fitting as a result of using small sample of each malware family [10,11].

The increasing rate at which machine learning and deep learning models fails to predict previously unseeing out-of-distribution zero-day malware despite state-of-the-art research and development on few-shot learning put a big question mark on the effectiveness of current state-of-the-art few-shot learning (FSL) approaches, and hence the importance and necessity of our investigation. In this research, we did an investigation into four (4) state-of-the-art few-shot learning to research their level of efficiency in detecting previously unseen malware variant or family. We started by pulling down the original research artifact comprising of the source code, data, etc, replicating the experiment with the original dataset, repeating the experiment with new malware dataset containing most recent malware variant that are not in existence when the original research was carried out, finally computing and comparing the MIN-MAX validation loss between training with the original dataset and new dataset containing most recent malware variants that does not previously exists. Our investigation aims the following contribution,

- To measure the efficiency of the current state-of-the-art few-shot learning against out-of-distribution zero-day malware.
- Depending on the result, to determine whether a new regularization as an improvement to current state-of-the-art few-shot learning approach or if a new approach is needed to match the frequency at which new malware variants are being developed and release
- To analyse variations in malware behavior relative to their success in fooling model to evade detection

2. Related Work

One major challenge on both machine learning-based and Deep Learning-based model is data scarcity, and this directly impacts performance of these models in a proportionate manner i.e A large amount of data is needed to achieve exceptional performance from a Deep learning model [12]. DL models are extremely data-hungry models because they needed a huge amount of labelled data to automatically learnt data-representation by themselves. Unfortunately, these data are not available thereby creating a significant challenge as their scarcity have direct implication on the performance of DL models. In order to ensure that exceptional performance could be obtained from Deep learning models in the presence of data scarcity, several research work had been done, in this section, we will look at recent state-of-the-art research on Single-Shot Learning (SSL), Few-Shot Learning (FSL), and Zero-Shot Learning (ZSL).

2.1. Single-Shot Learning (SSL)

To address problem of data scarcity or scantiness, One-shot learning uses a conceptualized approach whereby machine learning and deep learning models learns from only a single sample from each represented class thereby given the model capability to recognize and generalize patterns based on previously seen single example [13–15] In deep learning, it is not unusual to use discriminative embeddings or generative models as an alternative way for one-shot learning in the presence of data scarcity, but while they may be plausible for some classification tasks, the fact that they requires large amount of data makes them unsuitable for one-shot learning. Bertinetto et al. [16] proposed the learning of parameters of deep-learning in one-shot by constructing a second deep network called learnet which has the capability to predict pupil network parameter from a single sample, hence were able to obtained a forward one-shot learner which minimizes the one-shot objective through an end-to-end training.

Anton et al. [17] proposed a Deep Reinforcement One-shot Learning (DeROL) framework by training a deep-Q network with the sole aim of achieving policy that would be oblivious to unseen classes in the validation set, then each state of the one-shot learning process is further map to operation actions based on the trained deep-Q network which aids the maximization of the objective function

2.2. Few-Shot Learning (FSL)

Unlike Single-shot, few-shot learning (FSL) framework aims to address problems of data scarcity or scantiness by using few samples from each class label[18–20]. It leverages a large number of similar tasks so as to adapt a base-learner to a new task for which only few samples are available. Considering the fact that this framework uses few samples and deep learning models tends to overfit in the absence of huge amount of data, meta-learning uses of shallow neural network (SNN) to address the problem of overfitting that might arise from few samples. Qianru et al. [21] proposed meta-transfer learning (MTL) as a few-shot learning method whereby the model learns to adapt a neural network to a few-shot learning task by conducting experiments with (5-class, 1-shot) and (5-class, 5-shot) recognition tasks on two different challenging few-shot learning tasks, the proposed few-shot learning is a meta-learning based whereby the deep model is trained on multiple tasks and the learning is transfered by shifting, customizing and scaling functions of DNN weights to suit individual tasks.

Based on the intuition that some internal representations are more transferable than the other, Chelsea et al. [22] proposed a few-shot learning method that can learn the parameters of any model trained with gradient descent in a way that prepares the model for a fast adaptation such that it can learn new task from few samples. For model that is agnostic, the parameters are trained in a way that small amount of gradients steps coupled with a small amount of training sets from previously unseen tasks will produce good generalization for the task.

2.3. Zero-Shot Learning (ZSL)

In Zero-shot learning (ZSL), DNN model is trained both to recognize and categorize unseen objects. During validation, the learner is made to observe and predict samples from classes which were not previous observed during training to the class they belong to [23–25]. It uses some form of auxiliary information to associate observed and non-observed classes thereby making it possible to encode observable properties that distinguish an object from another, for instance, a previously trained DNN model to recognize a horse but not seen a zebra during training can still recognize both horse and zebra through their differences i.e the previously encoded differences. Considering that previously seen train set and unseen test validation set are mutually exclusive, zero-shot learning first map a relationship between unseen and seen class and then use the relationship to determine an unseen class during validation

Vinay et al. [26] proposed a generative framework for generalized zero-shot learning with a disjointed training and test classes through a feedback-driven mechanism in which the discriminator called multivariate regressor learns to map the generated exemplars to the corresponding class attribute vectors to create an improved generator. The proposed framework is variational autoencoder based architecture having a probabilistic encoder and emphconditional decoder thereby given the ability to generate samples from seen and unseen classes using each class attributes. The newly generated exemplar is then used to train any classification model.

3. Research Methodology

3.1. Dataset

3.1.1. Malimg Dataset

Being a publicly available dataset in Kaggle, Malimg dataset [27] contains a total number of 9435 executable malwares taken from 25 malware families which were previously disarmed before being converted to 32 by 32 images based on the nearest neighbor interpolation. Each Malware family in the dataset was shaped by transforming their binaries into matrix as a result of the conversion of malware binaries to 8-bit vectors leading to a 2D matrix of malware images

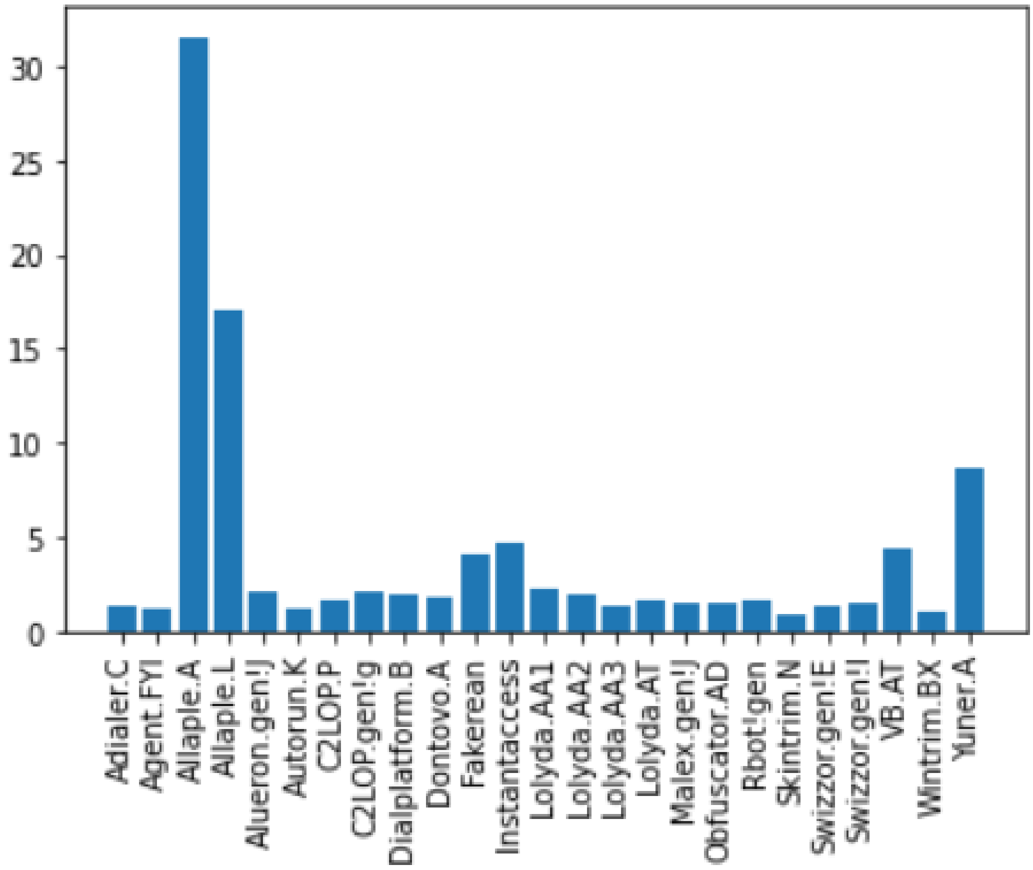

Figure 2. Malware Family Distribution in Malimg Dataset

3.1.2. Malevis Dataset

Malevis dataset consist of 26 malware family out of which one family represent "benign" or "legitimate" samples while the remaining 25 classes consist of different families of malware. The original binary images had been previously extracted from the malware files in 3 channels of RGB format before being resized into 224 by 224 and 300 by 300 dimension pixels while retaining the original number of channels in RGB format. Malevis dataset contains a total of 14,226 malware samples spanning 26 families of malware, and out of which 9100 are training samples while are 5126 validation samples in 3 channels format, the fact that the dataset makes provision for fairly larger legitimate malware samples for validation purpose makes the dataset suitable for the experiment

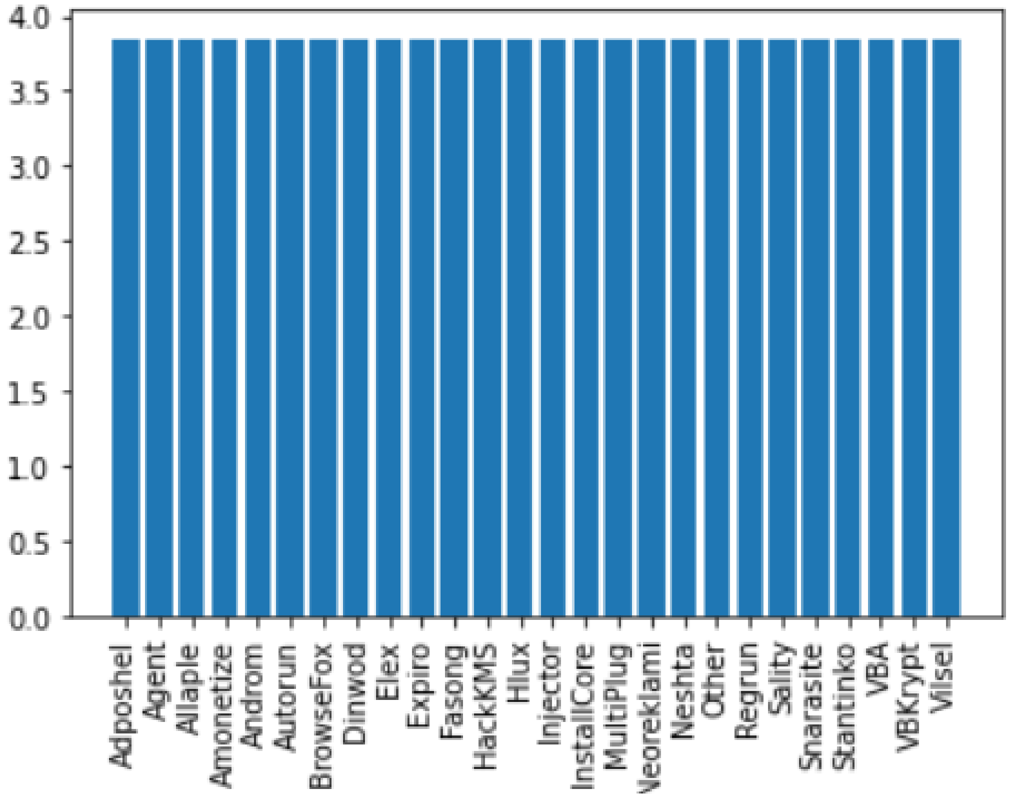

Figure 3. Malware Family Distribution in Malevis Dataset

3.2. Experimental Set-Up

Considering the imbalance across the distribution of malware family represented in malimg malware as compared with the relative balanced distribution of malware family samples in malevis dataset , choosing malevis malware dataset for the in-distribution becomes logical and reasonable while randomly selected malwares samples were chosen from malimg dataset for out-of-distribution. The experiment is in three (3) stages namely: zero-shot, single-shot, and few-shot. For each of the experimental-setup, malware samples from malevis dataset were used as in-distribution while randomly selected samples from malimg dataset were used as previously unseen out-of-distribution throughout the experiment.

3.2.1. Experimental Set-up with Zero-Shot Malware Samples

Considering that the two malware datasets are in different different channel with malimg in single channel while malevis was in three (3) channel of RGB. Substantial lines of python codes was written to automated the conversion of each and every samples in each represented malware families in malavis dataset from 3 channel to gray-scale single channel while also resizing them at the same time.

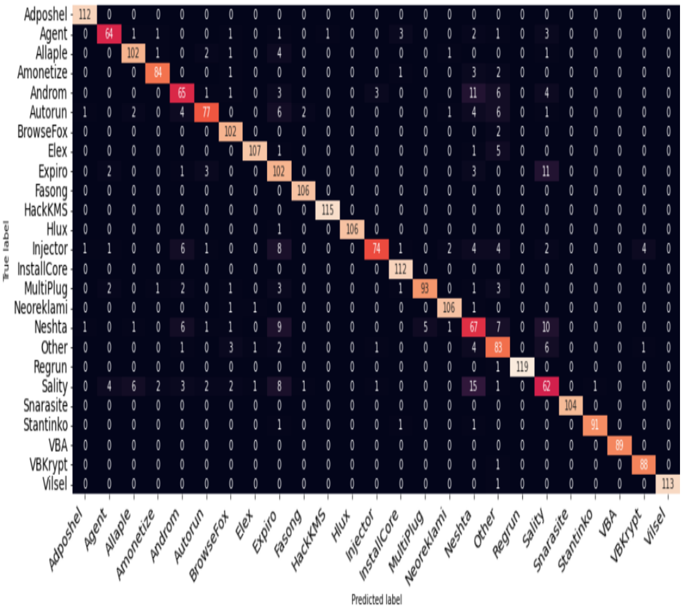

Figure 4. Confusion Matrix Resulting from Zero-shot malware experimental result brutally failed to classify any of the out-of-distribution (Dinwod) samples correctly despite awareness of the existence of the family during training

In order to investigate the performance of deep learning model on zero-shot, Dinwod malware family was removed from the training set, and then placed in the folder containing the validation set while ensuring that the removed malware family in the csv file contains the previously removed family name as part of the list of represented malware family in the distribution. This ensures that during training, the model is aware of of the existence of a malware family called Dinwod but without seen any of the samples from the Dinwod family, this technically made Dinwod family an out-of-distribution malware family.

That the model was not intelligent enough to classify and of the out-of-distribution (Dinwod) samples despite the awareness of such class during training leaves a gap to be explored. If the deep learning model could perform well base on previously seen samples during training, we assert that the model should be smart enough to detect and classify previously unseen data into the empty family as out-of-distribution considering that the model is made to be aware of the existence of such malware family during training.

3.2.2. Experimental Set-up with Single-Shot Malware Samples

In an era of data scarcity especially on new malware families were there is abundant of scarcity of the malware samples at the early stage of newly discovered family. We assert that state-of-the-art model should be able to detect and improve on its accuracy as more sample of newly discovered malware family samples becomes more available, this brings about the imperability of experimenting with a single representative sample for a whole malware family. Our approach to this was that, the model only sees a single out-of-distribution (Dinwod) sample during training but sees more of the out-of-distribution (Dinwod) samples during validation.

doi:10.20944/preprints202409.0509.v1

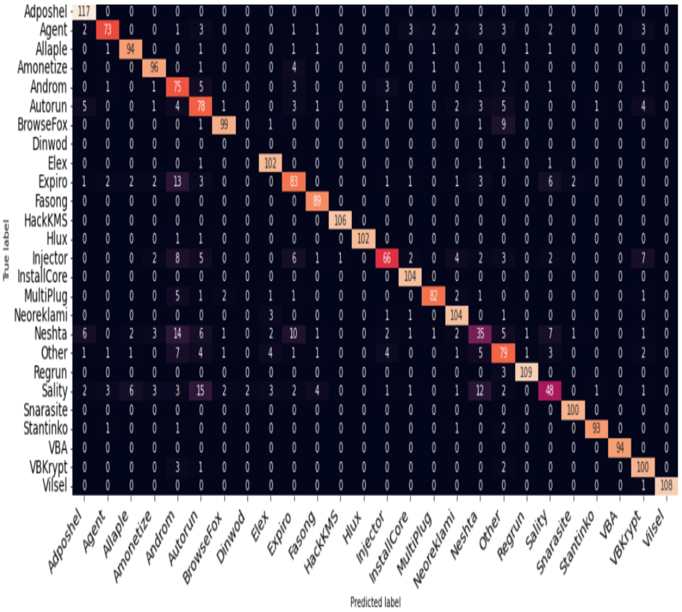

Figure 5. we expected Dinwod validation sample result to be ≥ 1 having previously seen one sample during training but not a single samples from the Dinwod out-of-distribution malware validation samples was was correctly predicted.

For single-shot, all previous processes taken on zero-shot was repeated but difference only in the sense that only a single out-of-distribution (Dinwod family) sample was introduced to the training sample and properly labelled accordingly. Unlike result obtained in single-shot where the model couldn't recognize the presence of any out-of-distribution family due to the the fact that none of the sample was previously seen during training, result obtained from single-shot was different in that the Dinwod malware family which represent our out-of-distribution sample was clearly seen in the confusion matrix, we asserted this to be as a result of the single Dinwod out-of-distribution malware family sample introduced to the training set, but none of the several out-of-distribution samples in the validation set was correctly predicted. Having previously seen one sample in training, model should be smart enough to predict some out-of-distribution correction. Hence we expected Dinwod validation sample result to be ≥ 1 but not a single sample from the Dinwod out-of-distribution malware validation samples was was correctly predicted.

3.2.3. Experimental Set-up with Few-Shot Malware Samples

To experiment with few shot, each steps taken during the experimentation with zero-shot and single-shot was repeated only that the number of out-of-distribution Dinwod malware family sample was increase from one to five to make it few-shot. With the increment in the number of out-of-dstribution malware samples introduced to the training set, it was expected that the correctly predicted out-of-distribution ≥ 1

doi:10.20944/preprints202409.0509.v1

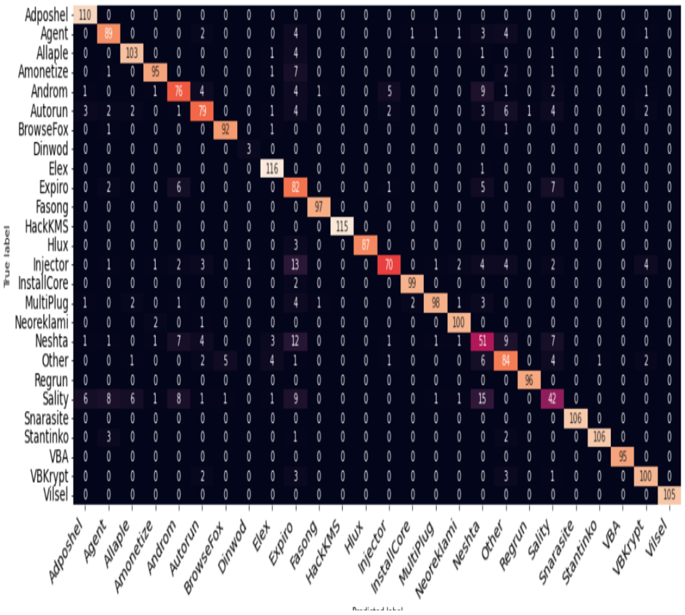

Figure 6. We asserted that deep learning model should be intelligent enough to detect an out-of-distribution malware sample having previously seen few of the samples during training.

due to increment in the number of out-of-distribution sample introduced to the validation set as the model was able to see and learn from more than one out-of-distribution Dinwod malware family samples during training. Albeit, validation result shows slight improvement in the overall accuracy but similar to the previous result obtain with single-shot in the sense that the correctly predicted out-of-distribution Dinwod malware family remains zero despite that the model was previously seen during training. We asserted that deep learning model should be intelligent enough to detect an out-of-distribution malware sample having previously seen few of the samples during training.

4. Conclusion

In this research, we exposed limitation and vulnerabilities of current state-of-the-art N-Shot (Zero-shot, Single-Shot, and Few-Shot) approaches to an out-of-distribution malware attack by doing an in-depth investigation into the performance of state-of-the-art Zero-shot, Single-shot, and few-shot learning approaches on zero-day out-of-distribution malware attack detection based on their static properties. During each experiment, we varied the number of samples in the out-of-distribution class accordingly to demonstrate zero-shot(no sample), single-shot (1 sample), few-shot(5 samples) while using confusion matrix to get the actual number of correct prediction on out-of-distribution malware validation samples. Result shows 0, 0, and 3 number of correct out-of-distribution predictions on Zero-shot, single-shot, and few-shot experiments respectively, thereby showing limitation on the current state-of-the-art approaches and the need for a more robust approach for zero-shot, single-shot, and few-shot as a model out-of-distribution attack detection.

We assert the plausibility of dynamic exploitation of the distribution of dimensional spaces between known malware families based on the combined Bayesian based algorithm and Deep learning classifier as a future research direction to improve current approaches for an effective detection of previously unseen out-of-distribution malware attack.

References

1. Ye, Y.; Li, T.; Adjeroh, D.; Iyengar, S.S. A survey on malware detection using data mining techniques. *ACM Computing Surveys (CSUR)* **2017**, *50*, 1–40.
2. Vasan, D.; Alazab, M.; Wassan, S.; Safaei, B.; Zheng, Q. Image-Based malware classification using ensemble of CNN architectures (IMCEC). *Computers & Security* **2020**, *92*, 101748.

3. Wang, P.; Tang, Z.; Wang, J. A novel few-shot malware classification approach for unknown family recognition with multi-prototype modeling. *Computers & Security* **2021**, *106*, 102273.
4. Ige, T.; Sikiru, A. Implementation of data mining on a secure cloud computing over a web API using supervised machine learning algorithm. Computer Science On-line Conference. Springer, 2022, pp. 203–210.
5. Okomayin, A.; Ige, T. Ambient Technology & Intelligence. *arXiv preprint arXiv:2305.10726* **2023**.
6. Pillai, S.E.V.S.; Polimetla, K. Mitigating DDoS Attacks using SDN-based Network Security Measures. 2024 International Conference on Integrated Circuits and Communication Systems (ICICACS). IEEE, 2024, pp. 1–7.
7. Vallabhaneni, R.; Vaddadi, S.A.; Pillai, S.; Addula, S.R.; Ananthan, B. Detection of cyberattacks using bidirectional generative adversarial network. *Indonesian Journal of Electrical Engineering and Computer Science* **2024**, *35*, 1653–1660.
8. Ige, T.; Marfo, W.; Tonkinson, J.; Adewale, S.; Matti, B.H. Adversarial sampling for fairness testing in deep neural network. *arXiv preprint arXiv:2303.02874* **2023**.
9. Ige, T.; Kiekintveld, C. Performance comparison and implementation of bayesian variants for network intrusion detection. 2023 IEEE International Conference on Artificial Intelligence, Blockchain, and Internet of Things (AIBThings). IEEE, 2023, pp. 1–5.
10. Ige, T.; Kiekintveld, C.; Piplai, A. An investigation into the performances of the state-of-the-art machine learning approaches for various cyber-attack detection: A survey. *arXiv preprint arXiv:2402.17045* **2024**.
11. Ige, T.; Kiekintveld, C.; Piplai, A. Deep Learning-Based Speech and Vision Synthesis to Improve Phishing Attack Detection through a Multi-layer Adaptive Framework. *arXiv preprint arXiv:2402.17249* **2024**.
12. Song, Y.; Wang, T.; Cai, P.; Mondal, S.K.; Sahoo, J.P. A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. *ACM Computing Surveys* **2023**, *55*, 1–40.
13. Wang, S.; Xu, M.; Sun, Y.; Jiang, G.; Weng, Y.; Liu, X.; Zhao, G.; Fan, H.; Li, J.; Zou, C.; others. Improved single shot detection using DenseNet for tiny target detection. *Concurrency and Computation: Practice and Experience* **2023**, *35*, e7491.
14. Zhu, W.; Zhang, H.; Eastwood, J.; Qi, X.; Jia, J.; Cao, Y. Concrete crack detection using lightweight attention feature fusion single shot multibox detector. *Knowledge-Based Systems* **2023**, *261*, 110216.
15. Lew, A.J.; Buehler, M.J. Single-shot forward and inverse hierarchical architected materials design for nonlinear mechanical properties using an Attention-Diffusion model. *Materials Today* **2023**, *64*, 10–20.
16. Bertinetto, L.; Henriques, J.F.; Valmadre, J.; Torr, P.; Vedaldi, A. Learning feed-forward one-shot learners. *Advances in neural information processing systems* **2016**, *29*.
17. Puzanov, A.; Zhang, S.; Cohen, K. Deep reinforcement one-shot learning for artificially intelligent classification in expert aided systems. *Engineering Applications of Artificial Intelligence* **2020**, *91*, 103589.
18. Jeong, J.; Zou, Y.; Kim, T.; Zhang, D.; Ravichandran, A.; Dabeer, O. Winclip: Zero-/few-shot anomaly classification and segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 19606–19616.
19. Dooley, S.; Khurana, G.S.; Mohapatra, C.; Naidu, S.V.; White, C. Forecastpfn: Synthetically-trained zero-shot forecasting. *Advances in Neural Information Processing Systems* **2024**, *36*.
20. Luo, X.; Wu, H.; Zhang, J.; Gao, L.; Xu, J.; Song, J. A closer look at few-shot classification again. International Conference on Machine Learning. PMLR, 2023, pp. 23103–23123.
21. Sun, Q.; Liu, Y.; Chua, T.S.; Schiele, B. Meta-transfer learning for few-shot learning. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 403–412.
22. Finn, C.; Abbeel, P.; Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. International conference on machine learning. PMLR, 2017, pp. 1126–1135.
23. Wang, Q.; Liu, L.; Jing, C.; Chen, H.; Liang, G.; Wang, P.; Shen, C. Learning conditional attributes for compositional zero-shot learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 11197–11206.
24. Liu, M.; Li, F.; Zhang, C.; Wei, Y.; Bai, H.; Zhao, Y. Progressive semantic-visual mutual adaption for generalized zero-shot learning. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 15337–15346.
25. Guo, J.; Guo, S.; Zhou, Q.; Liu, Z.; Lu, X.; Huo, F. Graph knows unknowns: Reformulate zero-shot learning as sample-level graph recognition. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, Vol. 37, pp. 7775–7783.

26. Verma, V.K.; Arora, G.; Mishra, A.; Rai, P. Generalized zero-shot learning via synthesized examples. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4281–4289.
27. Nataraj, L.; Karthikeyan, S.; Jacob, G.; Manjunath, B.S. Malware images: visualization and automatic classification. Proceedings of the 8th international symposium on visualization for cyber security, 2011, pp. 1–7.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.